

The invention claimed is:

1. A computer implemented method for transforming a plurality of chromatographic elution profiles, wherein each profile is obtained from the separation of a DNA mixture by Denaturing Matched Ion Polynucleotide Chromatography, wherein each DNA mixture comprises homoduplex and heteroduplex molecules obtained from the hybridization of a sample DNA and its corresponding wild type DNA, said method comprising:
  - a) overlaying said profiles on a coordinate system comprising a first axis associated with time values and a second axis associated with detector response values,
  - b) selecting first and second time points defining a time span wherein peaks due to said homoduplex and heteroduplex molecules are located within said span,
  - c) for each profile and within said span, adjusting the baseline by applying a slope factor to each detector response value, said factor derived from a line connecting the detector response values at said first and second time points, such that all of the profiles share a common baseline,
  - d) for each profile and within said span, normalizing the heights of the peaks to a pre-selected scale based on the height of the highest peak,
  - e) shifting the profiles along said first axis such that all of the profiles intersect at a pre-selected point on the last eluting peak of each profile within said span.
2. The method of Claim 1 wherein said pre-selected value is zero, wherein said pre-selected scale is from 0 to 1, wherein said pre-selected point comprises a point on the last eluting edge of said last eluting peak, and wherein in step (c) the second axis value at the first time point and the second axis value at the second time point are set to zero.
3. The method of Claim 1 wherein said profiles include at least one reference profile obtained from a standard mixture comprising the hybridization

product of DNA having a known sequence and corresponding wild type DNA.

4. A method for estimating the number of different single nucleotide polymorphisms in a plurality of same length DNA fragments, said method comprising:

5

a) hybridizing each of said same length DNA fragments with corresponding wild type DNA to form homoduplex and heteroduplex molecules,

10

b) analyzing the hybridization product of each of said same length DNA fragments by Denaturing Matched Ion Polynucleotide Chromatography to obtain a plurality of elution profiles,

c) transforming said profiles by the method of Claim 1,

15

d) sorting said transformed profiles into groups based on the shapes of said transformed profiles, wherein the number of single nucleotide polymorphisms is at least the same as the number of said groups.

5. A method for detecting the presence of a previously unknown single nucleotide polymorphism in a test DNA fragment, said method comprising:

20

a) hybridizing said test DNA fragment with corresponding wild type DNA

b) analyzing the product of step (a) by Denaturing Matched Ion

Polynucleotide Chromatography to obtain a test elution profile,

c) hybridizing a standard mixture comprising DNA fragments of known sequence with said wild type DNA,

d) analyzing the product of step (c) by Denaturing Matched Ion Polynucleotide Chromatography to obtain reference elution profiles,

25

e) obtaining a plurality of profiles by combining said test elution profiles and said reference elution profiles,

f) transforming said plurality of profiles by the method of Claim 1

g) sorting said plurality of profiles into groups based on the shapes of said plurality of elution profiles,

30

h) after said transforming, comparing said test elution profile with said groups, wherein said test DNA fragment is considered to contain a

previously unknown mutation if the shape of said test elution profile does not match the shapes of the profiles in said groups.

6. The method of Claim 5 further including subjecting said test DNA fragment to sequencing.

5 7. The method of Claim 1 including applying one or more statistical criteria to the transformed profiles obtained after step (e) to determine whether or not to group said transformed profiles into a single group.

8. The method of Claim 7 wherein said statistical criteria comprise:

10 a) within said span, dividing the first axis into a series of adjacent and evenly-spaced time regions wherein boundary lines, perpendicular to said first axis, are defined between adjacent regions, and wherein said profiles intersect said boundary lines at intersecting detector response values, b) for each boundary line

15 i) obtaining the mean of the intersecting detector response values, and comparing said mean to a first pre-selected value,

ii) obtaining the standard deviation of the mean of the intersecting detector response values, and comparing said standard deviation to a second pre-selected value,

20 iii) obtaining the range of the intersecting detector response values, and comparing said range to a third pre-selected value.

9. A computer implemented method for grouping a plurality of transformed chromatographic elution profiles obtained by the method of Claim 1, said method for grouping comprising:

25 a) within said span, dividing the first axis into a series of adjacent and evenly-spaced time regions wherein boundary lines, perpendicular to said first axis, are located between adjacent time regions, wherein said profiles intersect said boundary lines,

30 b) for each boundary line and between the highest intersecting profile and the lowest intersecting profile, dividing said each boundary line into a plurality of equally spaced and adjacent segments,

c) for each boundary line, numbered 1 through i

- i) determining the number of profiles intersecting each of said segments,
  - ii) determining the segment having the highest number of intersecting profiles and determining the nearest segment having zero intersecting profiles,
  - iii) for each boundary line, assigning a numerical grouping factor of  $n^i$  to the profiles that have a second axis value greater than said segment having zero intersecting profiles and assigning a grouping factor of 1 to the remaining intersecting profiles, wherein  $n$  is an integer greater than 1,
  - d) for each profile, obtaining a total value comprising the sum of all the grouping factors assigned to said each profile,
  - e) grouping together those profiles having the same total value.
10. The method of Claim 9 wherein  $n=2$ .
11. A method for grouping a plurality of transformed chromatographic elution profiles obtained by the method of Claim 1, said method for grouping comprising:
  - a) placing one or more markers, numbered 1 through  $i$ , each marker placed at a position where said transformed elution profiles show apparently clustered detector response values,
  - b) obtaining the first axis value and second axis value for each marker, each marker located on a boundary line perpendicular to said first axis,
  - c) for each marker, and along its associated boundary line, assigning a numerical grouping factor of  $n^i$  to the profiles that have a second axis value greater than the second axis value of said each marker, or otherwise assigning a grouping factor of 1 to the profiles, wherein  $n$  is an integer greater than 1,
  - d) for each profile, obtaining a total value comprising the sum of all the grouping factors assigned to said each profile,
  - e) grouping together those profiles having the same total value.
12. The method of Claim 11 wherein  $n=2$ .

13. A system for transforming chromatographic elution profiles, said system comprising:
- a computer having a processor and memory, wherein the computer receives a set of data corresponding to a plurality of chromatographic elution profiles, wherein each profile is obtained from the separation of a DNA mixture by Denaturing Matched Ion Polynucleotide Chromatography, wherein each DNA mixture comprises homoduplex and heteroduplex molecules obtained from the hybridization of a sample DNA and its corresponding wild type DNA, and wherein the processor:
- a) overlays said profiles on a coordinate system comprising a first axis associated with time values and a second axis associated with detector response values,
  - b) selects first and second time points defining a time span wherein peaks due to said homoduplex and heteroduplex molecules are located within said span,
  - c) for each profile and within said span, adjusts the baseline by applying a slope factor to each detector response value, said factor derived from a line connecting the detector response values at said first and second time points, such that all of the profiles have a common baseline,
  - d) for each profile and within said span, normalizes the heights of the peaks to a pre-selected scale based on the height of the highest peak,
  - e) shifts of the profiles along said first axis such that all of the profiles intersect at a pre-selected point on the last eluting peak of each profile within said span.
14. The system of Claim 13 wherein said pre-selected value is zero, wherein said pre-selected scale is from 0 to 1, wherein said pre-selected point comprises a point on the last eluting edge of said last eluting peak, and wherein in step (c) the second axis value at the first time point and the second axis value at the second time point are set to zero.
15. The system of Claim 13 wherein said processor:

f) within said span, divides the first axis into a series of adjacent and evenly-spaced time regions wherein boundary lines, perpendicular to said first axis, are located between adjacent time regions,

g) divides each boundary line into a plurality of equal and adjacent segments,

h) for each boundary line, numbered 1 through i

i) determines the number of profiles intersecting each of said segments,

ii) determines the segment having the highest number of intersecting profiles and determines the nearest segment having zero intersecting profiles,

iii) for each boundary line, assigns a numerical grouping factor of  $n^i$  to the profiles that have a second axis value greater than said segment having zero intersecting profiles and assigns a grouping factor of 1 to the remaining intersecting profiles, wherein n is an integer greater than 1,

i) for each profile, obtains a total value comprising the sum of all the grouping factors assigned to said each profile,

j) groups together those profiles having the same total value.

16. The method of Claim 15 wherein  $n=2$ .

17. The system of Claim 15 wherein said processor

f) receives instructions for placing one or more markers, numbered 1 through i, each marker placed at a position where said transformed elution profiles show apparently clustered detector response values,

g) obtains the first axis value and second axis value for each marker, each marker located on a boundary line perpendicular to said first axis,

h) for each marker, and along its associated boundary line, assigns a numerical grouping factor of  $n^i$  to the profiles that have a second axis value greater than the second axis value of said each marker, or otherwise assigns a grouping factor of 1 to the profiles, wherein n is an integer greater than 1,

- i) for each profile, obtains a total value comprising the sum of all the grouping factors assigned to said each profile,
- j) groups together those profiles having the same total value.

- 5            18. The system of Claim 17 wherein said processor applies one or more statistical criteria to the transformed profiles obtained after step (e) to determine whether or not to group said transformed profiles into a single group.
- 10           19. A computer readable medium for storing computer readable instructions, the instructions being capable of programming a computer to perform a method, the method comprising:
- 15                a method for transforming a plurality of chromatographic elution profiles, wherein each profile is obtained from the separation of a DNA mixture by Denaturing Matched Ion Polynucleotide Chromatography, wherein each DNA mixture comprises
- 20                homoduplex and heteroduplex molecules obtained from the hybridization of a sample DNA and its corresponding wild type DNA, said method for transforming comprising:
- 25                a) overlaying said profiles on a coordinate system comprising a first axis showing time values and a second axis showing detector response values,
- 30                b) selecting first and second time points defining a time span wherein peaks due to said homoduplex and heteroduplex molecules are located within said span,
- c) for each profile and within said span, adjusting the baseline by applying a slope factor to each detector response value, said factor derived from a line connecting the detector response values at said first and second time points, such that all of the profiles have a common baseline and the second axis value at the first time point and the second axis value at the second time point are set to zero,
- d) for each profile and within said span, normalizing the heights of the peaks to a pre-selected scale based on the height of the highest peak,

e) shifting the profiles along said first axis such that all of the profiles intersect at a pre-selected point on the last eluting peak of each profile within said span.

5 20. A computer readable medium for storing computer readable instructions, the instructions being capable of programming a computer to perform a method, the method comprising:  
the method for transforming of claim 19 and further comprising applying one or more statistical criteria to the transformed profiles obtained after step (e) to determine whether or not to group said transformed profiles into  
10 a single group.

21. A computer readable medium for storing computer readable instructions, the instructions being capable of programming a computer to perform a method, the method comprising:  
the method for transforming of claim 19 and further comprising a method  
15 for grouping a plurality of transformed chromatographic elution profiles, said method for grouping comprising:  
a) within said span, dividing the first axis into a series of adjacent and evenly-spaced time regions wherein boundary lines, perpendicular to said first axis, are located between adjacent time regions, wherein said profiles  
20 intersect said boundary lines,  
b) for each boundary line and between the highest intersecting profile and the lowest intersecting profile, dividing said each boundary line into a plurality of equally spaced and adjacent segments,  
c) for each boundary line, numbered 1 through i  
25 i) determining the number of profiles intersecting each of said segments,  
ii) determining the segment having the highest number of intersecting profiles and determining the nearest segment having zero intersecting profiles,  
30 iii) for each boundary line, assigning a numerical grouping factor of  $n^i$  to the profiles that have a second axis value greater than said



- segment having zero intersecting profiles and assigning a grouping factor of 1 to the remaining intersecting profiles, wherein  $n$  is an integer greater than 1,
- d) for each profile, obtaining a total value comprising the sum of all the grouping factors assigned to said each profile,
- e) grouping together those profiles having the same total value.
22. The computer readable medium of Claim 20 wherein  $n=2$ .
23. A computer readable medium for storing computer readable instructions, the instructions being capable of programming a computer to perform a method, the method comprising:
- the method for transforming of claim 19 and further comprising a method for grouping a plurality of transformed chromatographic elution profiles, said method for grouping comprising:
- a) placing one or more markers, numbered 1 through  $i$ , each marker placed at a position where said transformed elution profiles show apparently clustered detector response values,
- b) obtaining the first axis value and second axis value for each marker, each marker located on a boundary line perpendicular to said first axis,
- c) for each marker, and along its associated boundary line, assigning a numerical grouping factor of  $n^i$  to the profiles that have a second axis value greater than the second axis value of said each marker, or otherwise assigning a grouping factor of 1 to the profiles, wherein  $n$  is an integer greater than 1,
- d) for each profile, obtaining a total value comprising the sum of all the grouping factors assigned to said each profile,
- e) grouping together those profiles having the same total value.
24. The computer readable medium of Claim 23 wherein  $n=2$ .
25. A plurality of transformed elution profiles obtained by the method of Claim 1.
26. A plurality of elution profiles grouped by the method of Claim 9.
27. A plurality of elution profiles grouped by the method of Claim 11.